



(19) Europäisches Patentamt  
European Patent Office  
Office européen des brevets



(11) EP 0 850 673 A1

(12)

**EUROPEAN PATENT APPLICATION**  
published in accordance with Art. 158(3) EPC

(43) Date of publication:  
01.07.1998 Bulletin 1998/27

(51) Int. Cl.<sup>6</sup>: A63F 9/22

(21) Application number: 97930773.3

(86) International application number:  
PCT/JP97/02421

(22) Date of filing: 11.07.1997

(87) International publication number:  
WO 98/02223 (22.01.1998 Gazette 1998/03)

(84) Designated Contracting States:  
DE ES FR GB IT

• HAMADA, Kazuhiko,  
Sega Enterprises, Ltd.  
Tokyo 144 (JP)

(30) Priority: 11.07.1996 JP 182495/96

(74) Representative:  
Brown, Kenneth Richard et al  
R.G.C. Jenkins & Co.  
26 Caxton Street  
London SW1H 0RJ (GB)

(71) Applicant:  
SEGA ENTERPRISES, LTD.  
Tokyo 144 (JP)

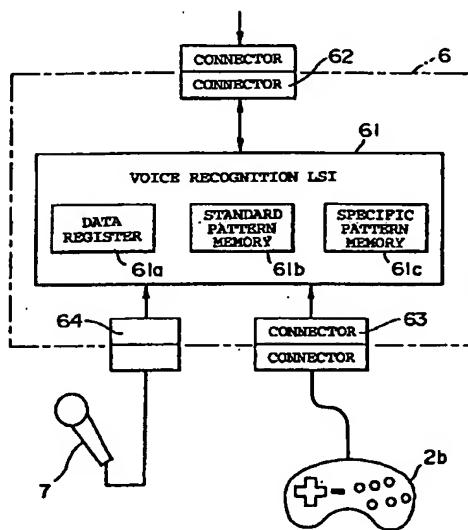
(72) Inventors:

• NAKAJIMA, Takuya,  
Sega Enterprises, Ltd.  
Tokyo 144 (JP)

**(54) VOICE RECOGNIZER, VOICE RECOGNIZING METHOD AND GAME MACHINE USING THEM**

(57) In order to play a game by voice control, a voice recognition device used as a peripheral device and a voice recognition method are provided. The voice recognition device comprises voice input means 7, a game machine control section 2b, and a voice recognition section 61 for recognizing the player's voice by comparing the voice signal output from this voice input means 7 with data from previously defined voice recognition dictionaries 61b, 61c, and generating control signals relating to the game on the basis of the recognition result and output signals from the control section 2b.

FIG. 3



**Description****TECHNICAL FIELD**

5 This invention relates to a voice recognition device and voice recognition method for recognizing input voices, and a game machine using same.

**BACKGROUND ART**

10 With the progress of computer technology in recent years, video game machines using computer graphics technology have become widely used. A wide variety of different game software has been developed recently, so domestic video game machines of this kind have become extremely popular as a form of entertainment. In a video game machine of this kind, characters are moved and the game screen is changed by operating buttons on a controller.

15 In a conventional video game device, all game instructions are given by means of the player operating controller switches. However, for humans, the most natural means of communication is the voice. Therefore, various attempts have been made to recognize the human voice and to control machines on the basis of these recognition results. For example, attempts have been made to control cars, robots, or the like by means of the human voice, or to control various electronic devices by vocal commands.

20 However, when it is sought to apply a conventional voice recognition device to a video game machine, the device must conform to the characteristics of the game machine, unlike the case of general electronic devices, and it must be adapted before application. For example, when the player presses a button on the game machine controller, rather than pressing the button just once, the degree of movement of the character can be changed by pressing the button for a long time or pressing it repeatedly, and the character can be made to perform special actions by pressing the button simultaneously with another button, for example. It is also important that the device has a good response. On the other 25 hand, in some types of game, the accuracy of recognition may not be particularly important, indeed, the fact that the device cannot recognize the player's voice very well may itself increase the enjoyment of the game.

In brief, for the player to enjoy the game, it is necessary to provide particular characteristics and functions which differ from those in a conventional voice recognition device. A voice recognition device which satisfies these demands is not known in the prior art.

30 This invention was devised in order to resolve the related problems, an object thereof being to provide a voice recognition device and voice recognition method suited for a game.

It is a further object of this invention to provide a game machine using this voice recognition device and voice recognition method.

**DISCLOSURE OF INVENTION**

35 In a voice recognition device used as a peripheral device for a game machine, the voice recognition device relating to this invention comprises voice input means, and a voice recognition section for recognizing the player's voice by comparing the voice signal output from this voice input means with data from previously defined voice recognition dictionaries, and generating control signals relating to the game on the basis of the recognition result.

In the voice recognition device relating to this invention, the voice recognition section comprises a non-specific speaker voice recognition dictionary, which is previously defined for unspecified speakers, and a specific speaker voice recognition dictionary which is defined by the player, and in its initial state, the device selects the non-specific speaker voice recognition dictionary.

40 In the voice recognition device relating to this invention, the voice recognition section comprises a plurality of specific speaker voice recognition dictionaries corresponding respectively to a plurality of players, and one of these specific speaker voice recognition dictionaries is selected by an action of the player as the dictionary to be used for voice recognition processing.

The voice recognition device relating to this invention comprises a game machine control section connected to the 45 voice recognition section, and the voice recognition section generates control signals relating to the game by combining voice recognition results and control signals from the control section.

In the voice recognition device relating to this invention, the control section outputs control signals for implementing normal actions, and the voice recognition section generates control signals for implementing special actions.

50 "Normal actions" are actions normally implemented by the player in the game (for example, kicking, jumping, etc.) "Special actions" are particular actions which are made possible by the combination of signals from the control section and voice recognition signals. For example, in a fighting game, by a combination of a raised voice + operation of button A, the player may implement an action whereby the same move is repeated, or a particular deathblow which can be produced only when a plurality of buttons are pressed simultaneously.

In the voice recognition device relating to this invention, the voice recognition section outputs a value indicating the state of the voice signal and a similarity level indicating the degree of similarity between the voice signal output from the voice input means and the contents of the voice recognition dictionary, and the voice recognition section has a first operating mode, wherein voice recognition is conducted on the basis of the similarity level in order to select the type of game action, and a second operating mode, wherein no voice recognition is conducted and the state of the voice signal is measured in order to set the state of the action.

In the second mode, the voice recognition device relating to this invention takes the average sound level of a part or all of the voice signal as the state of the voice signal, and sets the state of the game action on the basis of this average sound level.

10 In the second mode, the voice recognition device relating to this invention takes the peak sound level of the voice signal as the state of the voice signal, and sets the state of the game action on the basis of this peak sound level.

In the second mode, the voice recognition device relating to this invention takes the voice signal rise time as the state of the voice signal, and sets the state of the game action on the basis of this rise time.

15 In the second mode, the voice recognition device relating to this invention takes the voice signal continuation time as the state of the voice signal, and sets the state of the game action on the basis of this continuation time.

In the second mode, the voice recognition device relating to this invention takes the type of voice as the state of the voice signal, and sets the state of the game action on the basis of this voice type.

20 In the voice recognition device relating to this invention, the voice recognition section outputs a similarity level indicating the degree of similarity between the voice signal output from the voice input means and the contents of the voice recognition dictionary, and the corresponding volume level, evaluates this volume level on the basis of a predetermined rejection level, and rejects the recognition result from the voice recognition section depending on this evaluation result.

The voice recognition device relating to this invention sets a rejection level for each type of game or each game stage.

25 In the voice recognition device relating to this invention, the voice recognition section performs voice recognition on the basis of the similarity level indicating the degree of similarity between the voice signal output from the voice input means and the contents of the voice recognition dictionary, and changes the state of action in the game according to control signals generated on the basis of this recognition result, in response to this similarity level.

The game machine relating to this invention comprises the voice recognition device as a peripheral control device.

30 The voice recognition method relating to this invention comprises: a first step whereby a voice signal is received; a second step whereby the player's voice is recognized by comparing this voice signal with data from a previously defined voice recognition dictionary; and a third step whereby control signals relating to the game are generated on the basis of the recognition result from the second step.

35 The voice recognition method relating to this invention comprises a fourth step whereby control signals are received from the game machine control section, and in the third step, control signals relating to the game are generated by combining the voice recognition result from the second step and the control signals from the fourth step.

#### BRIEF DESCRIPTION OF DRAWINGS

Fig. 1 is an external view of a game system using a video game machine relating to a first mode for implementing this invention; Fig. 2 is a functional block diagram of a video game machine relating to a first mode for implementing this invention; Fig. 3 is a functional block diagram of a voice recognition section relating to a first mode for implementing this invention; Fig. 4 is an illustrative diagram of operational states of a game system using a video game machine relating to a first mode for implementing this invention; Fig. 5 is a flowchart for describing the action of a voice recognition section relating to a first mode for implementing this invention; Fig. 6 is an approximate diagram of a voice waveform for describing the action of a voice recognition section relating to a first mode for implementing this invention; Fig. 7 is an approximate diagram of a voice waveform for describing the action of a voice recognition section relating to a first mode for implementing this invention.

#### BEST MODE FOR CARRYING OUT THE INVENTION

##### First mode for implementing this invention

Fig. 1 is an external view of a video game machine using a voice recognition device relating to a mode for implementing this invention. In this diagram, the video game machine main unit 1 is approximately box-shaped and it contains circuit boards and the like for game processing. The front face of the video game machine main unit 1 is provided with two connectors 2a and a PAD 2b for controlling the game is connected to one of these connectors 2a via a cable 2c. The other connector 2a is connected to a voice recognition section 6. A PAD 2b and a microphone 7 for inputting the player's voice are connected to the voice recognition section 6. If two people are playing the game, then two PADs

2b are used. The microphone 7 shown in Fig. 1 is a headset microphone, but another type of microphone may be used. Furthermore, the voice recognition section 6 may be connected to both of the connectors 2a.

A cartridge I/F 1a and CD-ROM drive 1b are provided on the top of the video game machine main unit 1 for respectively connecting ROM cartridges and reading CD-ROMs, which serve as recording media on which game programs and voice recognition operations are recorded. Although omitted from the drawing, a video output terminal and audio output terminal are provided on the rear side of the video game machine main unit 1. The video output terminal is connected via a cable 4a to a video input terminal of a TV receiver 5, and the audio output terminal is connected via a cable 4b to an audio input terminal of the TV receiver 5. In a video game machine of this kind, the user can play a game by operating the PAD 2b whilst watching screens shown on the TV receiver 5.

Fig. 2 is a block diagram showing an overview of a TV game machine relating to this mode of implementation. The voice recognition section 6 and microphone 7 are not shown in Fig. 2. This image processing device comprises a CPU block 10 for controlling the whole device, a video block for controlling the display of game screens, a sound block 12 for generating sound effects, etc., and a subsystem 13 for reading CD-ROMs, and the like.

The CPU block 10 comprises an SCU (System Control Unit) 100, main CPU 101, RAM 102, ROM 103, cartridge I/F 1a, sub-CPU 104, CPU bus 103, and the like. The main CPU 101 controls the whole device. This main CPU 101 comprises an internal calculating function similar to a DSP (Digital Signal Processor) and is capable of implementing applicational software at high speed. The RAM 102 is used as a work area for the main CPU 101. The ROM 103 stores initial programs and the like which are used for initialization processing. The SCU 100 conducts smooth data input and output between the main CPU 101, VDPs 120, 130, DSP 140, CPU 141, etc., by controlling buses 105, 106, 107. The SCU 100 is provided with an internal DMA controller and is capable of transferring sprite data for the game to the VRAM in the video block 11. Thereby, applicational software for a game, or the like, can be implemented at high speed. The cartridge I/F 1a is used for inputting applicational software supplied in ROM cartridge format.

The sub-CPU 104 is known as an SMPC (System Manager & Peripheral Control), and it is provided with a function for collecting peripheral data from the PADS 2b via the connectors 2a in response to requests from the main CPU 101. The main CPU 101 conducts processing such as moving a character on the game screen, for example, on the basis of peripheral data read in from the sub-CPU 104. A desired type of peripheral, including a PAD, joystick, keyboard, or the like, can be connected to the connectors 2a. The sub-CPU 104 is provided with a function whereby it automatically identifies the type of peripheral connected to a connector 2a (main unit terminal), and collects peripheral data and the like by means of a communications format which corresponds to the peripheral type.

The video block 11 comprises a VDP (Video Display Processor) 120 which draws characters etc. consisting of polygon data for the video game, and a VDP 130 which draws background screens, synthesizes the polygon image data and background images, and performs clipping, and other processes. VDP 120 is connected to a VRAM 121 and frame buffers 122, 123. Polygon picture data representing characters in the video game machine is transferred from the main CPU 101 via the SCU 100 to VDP 120, where it is written into VRAM 121. The picture data written into VRAM 121 is then drawn into picture frame buffer 122 or 123 in a 16 or 8 bit/pixel format, for example. The data drawn into frame buffer 122 or 123 is transferred to VDP 130. The information for controlling the drawing process is supplied from the main CPU 101 via the SCU 100 to the VDP 120. The VDP 120 then implements drawing processes in accordance with these instructions.

The VDP 130 is connected to the VRAM 131 and is constructed such that picture data output from the VDP 130 is output to an encoder 160 via a memory 132. The encoder 160 generates a video signal by appending a synchronizing signal, or the like, to this image data, and outputs this signal to the TV receiver 5. Thereby, prescribed game screens are displayed on the TV receiver 5.

The sound block 12 comprises a DSP 140 which synthesizes sound by means of a PSM or FM system, and a CPU 141 which controls this DSP 140, and the like. Sound data generated by the DSP 140 is converted to a two-channel signal by the D/A converter 170 and then output to speakers 5b.

The sub-system 13 comprises a CD-ROM drive 1b, CD I/F 180, CPU 181, MPEG AUDIO 182, MPEG VIDEO 183, and the like. This sub-system 13 is provided with a function whereby it reads in applicational software supplied in CD-ROM format and reproduces animated pictures, or the like. The CD-ROM drive 1b reads in data from a CD-ROM. The CPU 181 conducts processes such as controlling the CD-ROM drive 1b and correcting errors in the input data. The data read from the CD-ROM is supplied via a CD I/F 180, bus 106, and SCU 100 to the main CPU 101, and is used as applicational software. The MPEG AUDIO 182 and MPEG VIDEO 183 are devices which restore data compressed according to MPEG (Motion Picture Expert Group) standards. By using the MPEG AUDIO 182 and MPEG VIDEO 183 to restore MPEG-compressed data written onto a CD-ROM, it is possible to reproduce animated pictures.

Fig. 3 is a block diagram showing the internal composition of a voice recognition section 6. The voice recognition section 6 is a voice recognition subsystem for individual words (words used independently, such as "sun" or "sea", etc.), which is provided with a word spotting function that corresponds to unspecified speakers and can recognize anyone's voice, rather than requiring the user to register his or her own voice. In the diagram, connector 62 is connected via cable 2c to connector 2a in Fig. 2. Furthermore, connectors 63 and 64 are connected respectively to PAD 2b and microphone

7. Connectors 62 - 64 are connected to a voice recognition LSI 61 (model: RL5C288 : manufactured by Ricoh Co. Ltd.). The voice recognition LSI 61 conducts voice recognition processing on the basis of voice signals input from the microphone 7, and it transfers data from PAD 2b to the main unit CPU block 10. Processing is also conducted to combine the control data from the PAD 2b with the voice recognition result. By connecting a compatible peripheral to the expansion port (connector 63), simultaneous voice and manual operation are possible. The voice recognition LSI 61 comprises a data register 61a for storing parameter data etc. for voice recognition processing, a standard pattern memory 61b for storing standard patterns for voice recognition, and a specific pattern memory 61c for voice recognition of a specific speaker.

The voice recognition section 6 previously classifies (by clusters) the spectral series of a plurality of speakers for each word, and it takes the centre of each cluster or the average spectral series of the voices belonging to each cluster as standard patterns (multi-templates). Input voices are matched with each standard pattern to calculate a similarity level. The similarity level is contained as data in the recognition result and it indicates the distance between the vocal sound made by the user and word resulting from the recognition process. The higher its value, the greater the similarity. A standard pattern word having a high similarity level is output as a voice recognition result. Standard pattern dictionary data is downloaded from the game machine 1. Furthermore, as well as the word similarity level, the received sound volume is also output as part of the recognition result. The received sound volume is a value indicating the level of the input voice, and by this means is it possible to measure the loudness of the vocal sound made by the user. The higher the value, the greater the volume.

The voice recognition section 6 has the following characteristics.  
 20 Recognition type: Individual spoken word; speaker type: unspecified speaker; recognition system: DST model; number of recognized words: 30 standard words for unspecified speaker (number of recognized words can change with the total number of moras); recognition processing time: 0.2 seconds or more from end of speaking (depends on applicational software); word length: 0.15 - 2.00 seconds (including meaningless words); clustering system: bit assignment system; number of clusters: maximum 16; rejection levels: 8 stages (0 (low level) - 7 (high level)); word connection: minimum 30 (number of combined words may change with number of moras), microphone amplifier gain: variable.

Fig. 4 is a compositional example of a system using this game machine 1 and voice recognition section 6. The details are described later, but the player (PLAYER) can implement various actions by operating the PAD 2b and giving commands via the microphone 7, whilst watching the screen of the TV receiver 5.

Fig. 5 is a flowchart of the basic processing in voice recognition. In the following description, "word" does not mean a word in a strict grammatical sense, but rather denotes a single unit used in voice recognition, including short items, such as a person's name or a letter, or long items, such as a phrase or sentence, etc.

Firstly, a recognition command is set in the data register 61a (step S10). The device then waits until the voice recognition LSI 61 outputs a recognition result (step S11).

Having received a recognition command, the voice recognition LSI 61 stores a recognition result in data register 35 61a when a voice of a prescribed volume level or above has continued for a prescribed period of time, and it informs the game machine main unit 10 that a recognition result has been output.

When a recognition result has been output, the word having the highest recognition ranking, which is regarded to be the most similar of the plurality of standard patterns, is read out from data register 61a, along with a recognition score (step S12). The recognition score expresses the similarity between the input voice and the word having the highest recognition ranking in the form of a points score. The higher the recognition score, the greater the resemblance to the input voice.

Next, it is determined whether or not more detailed data is required (step S13), and if data is not required, the procedure returns to step S10.

If data is required, a "next" command is set in data register 61a (step S14). In accordance with the "next" command, 45 the voice recognition LSI 61 stores volume data for the spoken recognized word in data register 61a. The game machine main unit 10 reads out the volume data from the data register 61a (step S15).

Next, it is determined whether or not more detailed data is required (step S16), and if data is not required, the procedure returns to step S10.

If data is required, a "next" command is set in data register 61a (step S17). In accordance with the "next" command, 50 the voice recognition LSI 61 stores the word having the second highest recognition ranking and the corresponding recognition score in the data register 61a. The game machine main unit 10 reads out this word having the second highest recognition ranking and its recognition score from the data register 61a (step S18).

Next, it is determined whether or not more detailed data is required (step S19), and if data is not required, the procedure returns to step S10.

If data is required, a "next" command is set in data register 61a (step S20). In accordance with the "next" command, 55 the voice recognition LSI 61 stores the word having the third highest recognition ranking and the corresponding recognition score in the data register 61a. The game machine main unit 10 reads out this word having the third highest recognition ranking and its recognition score from the data register 61a (step S21).

In this way, it is possible to obtain the word having the highest recognition ranking and its recognition score as a recognition result from the voice recognition LSI 61, and depending on requirements, it is also possible to obtain more detailed recognition result data by means of a "next" command.

By means of the voice recognition processing described above, the voice recognition section 6 has the following functions.

#### Default non-specific speaker voice recognition function

The voice recognition section 6 switches between a default recognition mode and normal recognition mode according to a prescribed sequence. The object of the default recognition mode is to enable initial screens for operating a CD-ROM etc. displayed immediately after the device power has been turned on to be operated by voice. An internal ROM of the voice recognition LSI 61, which is omitted from the diagram, contains a word dictionary corresponding to each button of the PAD 2b, and if a corresponding word is recognized, the bit corresponding to each button is set to active (0). In this mode, initialization settings immediately after switching on the power, or the start of a game using a CD-ROM, etc. can be activated by voice.

The relationship between the word dictionary and corresponding buttons is shown below.

#### Relationship between default word dictionary and corresponding buttons

20	Pad button	Word dictionary
	Right	None
	Left	None
	Down	None
	Start	"Start game"
25	A button	"Set"
	C button	"Button C"
	B button	"Delete"
	R button	"Forwards"
	X button	"Repeat"
30	Y button	"Stop"
	Z button	"Replay"
	L button	"Backwards"

#### Normal non-specific speaker voice recognition function

The object of this mode is to load a dictionary from a CD-ROM in the game machine 1 and to use it in a game by conducting voice recognition. A dictionary data download command is issued, a dictionary is loaded from the CD-ROM to a RAM in the voice recognition LSI 61, which is omitted from the diagram, and the voice recognition operation is started by issuing a recognition start command (the rejection level and cluster to be used must be set, as described below). The recognition operation can be halted by issuing a recognition halt command.

In this case, the device may be set such that the dictionary data is downloaded automatically when the power is turned on or the device is reset. If the power to the voice recognition section 6 is turned on or reset, then data from the CD-ROM is automatically transferred to the internal system RAM. Thereupon, substitution with standard pad data is carried out. If data download is conducted automatically in this way, then since no external operation is required, standard data can be used by the application even if the application used is incompatible. Furthermore, by providing standard data in the system in advance, the data volume for each type of application can be cut and the load on the creator of the application can be reduced.

#### Word spotting function

The voice recognition section 6 has a word spotting function whereby the word in the dictionary is isolated from a phrase containing unnecessary sounds such as the "Err" in "Err ...punch".

#### Catch up function

The voice recognition section 6 employs non-specific speaker voice recognition on the basis of the standard pattern memory 61b. However, in some cases, it may be difficult to recognise a user with particularly idiosyncratic pronunciation or specific words. Therefore, it is also possible to achieve improvements in the recognition rate by rewriting words that

are difficult to recognize in an independent user voice register. In this case, specific speaker voice recognition is used.

The concrete details of the procedure are as follows. A voice data transfer command is issued to the data register 61a. A desired voice sound is then input to the microphone 7 and the input voice sound is stored in a specific pattern memory 61c.

5

#### Microphone gain changing function

The microphone 7 gain can be changed. The gain can be set by issuing a gain setting command to the data register 61a. By means of this function, suitable voice recognition can be achieved with respect to the volume of the speaker's voice, peripheral noise, and the like.

10

#### Real time volume output function

The normal recognition mode comprises two recognition modes: a noise switch mode where the voice level is output in real time (every 20 ms), and a recognition mode where normal voice recognition is conducted. In the former, real-time volume output mode, since the voice level only is output, it is possible to conduct processing at extremely high speed. By issuing a mode volume output command, an 8-bit (0 - 255) voice level signal is output.

One conceivable application of this mode is as follows. After a command has been input from the microphone 7 or PAD 2b, the degree of the action carried out according to that command is determined from the voice level signal. For example, after a "forward" command has been given to a character on the screen, the player adds "quickly, quickly" in order to increase the character's speed of movement. Alternatively, if an expression, such as a tearful smile, or the like, is to be applied to the character on the screen, then the intensity of the expression is determined by the volume of the player's voice saying "Ahhh!", or the like. Alternatively, if the character on the screen is walking through dangerous terrain, then when a dangerous situation arises unexpectedly, for instance, an enemy character appears suddenly, by shouting "watch out!", "car!", or the like, the player can make the character slow down its speed of movement, adopt a defensive stance, lower its posture, fall unconscious, or the like, in response to the volume of the player's voice.

Rather than changing the degree of an action, it is also possible to change the type of action by means of the voice level. For example, after a "forward" command has been given to the character on the screen, if the player says "go" is a low voice, the character starts to walk, whereas if the player shouts "GO!", then the character will jump, or if it has a flying capability, it will fly through the sky.

In brief, it is possible for the degree of an action to be increased or reduced by means of the voice level, or for the type of action to be changed if the voice level exceeds a certain threshold value.

35

#### Reject function

This is a function whereby, during voice recognition, when the score (level) output by the voice recognition LSI 61 together with the similarity level is below a prescribed level, the corresponding recognition result is discarded. For example, if score < (rejection level + 1) \* 8, then the result having the highest similarity ranking will not be output. The reject level may be set to 2, for example. By means of this function, it is possible to prevent recognition errors due to voices other than that of the speaker and noise, etc.

#### Cluster function

This is a function whereby the recognition rate and recognition speed can be increased by dividing 30 words for a single scene into a maximum of 16 groups and then conducting voice recognition for a selected group (or plurality of groups).

For example, in order to recognise an incantation, or the like, which is used regularly in a fighting scene of a role playing game (RPG), then a cluster is created for each type of incantation, namely, magic for attacking (cluster 1), magic for recovering (cluster 2), direct attack (cluster 3), other (cluster 4), role (cluster 5), etc. Thereupon, as the fighting scene progresses, the appropriate clusters can be selected in sequence.

Before starting a fight, a word from the role cluster (cluster 5) is recognized. If the recognition result is "fighter", then the character will subsequently take a weapon and engage in a direct attack, so the voice recognition process moves to the direct attack cluster (cluster 3). On the other hand, if the recognition result for the role is "wizard", then the voice recognition process will be applied to the magic for attacking cluster (cluster 1).

In this way, an appropriate cluster which is closely related to the game scene is selected on the basis of the previous recognition result, and it is determined if the voice signal corresponds to any of the words in the selected cluster. By this method, there is little risk that that a completely unrelated word will be recognized in error, so the reliability of recognition is high, and since the words subjected to voice recognition for which a similarity is sought are limited to

## EP 0 850 673 A1

those in the selected cluster, the number of processing calculations is reduced, which is advantageous in terms of processing speed.

### Word connection function

5 This is function whereby a single recognition result is output for two words spoken consecutively. A result is output in this way when the sum of the similarity level for the first word and the similarity level for the second word is high. In this case, it is necessary to store connection data for the two words previously in the dictionary.

10 For example, the words "monkey", "attack", "pheasant", "defend", "dog", "magic", "devil", "escape", "momotaro" may be contained in the dictionary, and word connection data such as the following may be stored: "attack - monkey", "defend - monkey", "escape-monkey", "attack-pheasant", "defend-pheasant", "escape - pheasant", "attack - dog", "defend - dog" and "escape - dog". In this way, words and corresponding connection data are stored in the dictionary. "momotaro" is a character from a Japanese fairy tale.

15 In this case, if the user were to say "attack-monkey" (leaving a slight pause inbetween), then the action of the voice recognition section 6 might correspond to the following cases, for example (sequence in which three words are selected from nine words \* 2).

#### Case 1

20

1 <sup>st</sup> word similarity	2 <sup>nd</sup> word similarity	Recognition result
Rank 1 attack	Rank 1 monkey	
Rank 2 magic	Rank 2 pheasant	
Rank 3 escape	Rank 3 devil	Attack - monkey

#### Case 2

30

1 <sup>st</sup> word similarity	2 <sup>nd</sup> word similarity	Recognition result
Rank 1 attack	Rank 1 pheasant	
Rank 2 magic	Rank 2 monkey	
Rank 3 escape	Rank 3 devil	Attack - pheasant

40 Case 3

45

1 <sup>st</sup> word similarity	2 <sup>nd</sup> word similarity	Recognition result
Rank 1 attack	Rank 1 devil	
Rank 2 magic	Rank 2 monkey	
Rank 3 escape	Rank 3 dog	Attack - monkey

50

#### Case 4

55

1 <sup>st</sup> word similarity	2 <sup>nd</sup> word similarity	Recognition result
Rank 1 magic	Rank 1 devil	

(continued)

5

1 <sup>st</sup> word similarity	2 <sup>nd</sup> word similarity	Recognition result
Rank 2 attack	Rank 2 monkey	Attack - monkey
Rank 3 escape	Rank 3 dog	

When the aforementioned word connection function is used, the voice recognition section 6 selects an output from three candidates having the 1<sup>st</sup> - 3<sup>rd</sup> highest similarity rankings. The words having the 1<sup>st</sup>, 2<sup>nd</sup> and 3<sup>rd</sup> highest similarity rankings for the first word are combined with the words having the 1<sup>st</sup>, 2<sup>nd</sup> and 3<sup>rd</sup> highest similarity rankings for the second word, and the words having the highest total similarity level are selected. However, if, as in cases 3 and 4 above, a word for which there is no connection data comes highest in the similarity ranking, then this is discarded and a word for which there is connection data is given priority and is output.

Next, the operation of a game machine using a voice recognition section is described.

A game machine of this type is normally controlled by buttons on a pad 2b, but instead of this, or in addition to this, it is possible to use a new operating method by applying a voice recognition device. If the cause for any difference in volume can be identified, then this can be combined with voice recognition to provide a new way of playing a game.

In normal voice recognition, it is simply determined whether or not the vocal sound made by the player matches recognition data and the result of this judgement is output, but by appending to this judgement results relating to whether the voice is louder or softer than a certain standard level, it is possible to broaden the scope of application. A concrete example of this is described later.

The voice sound produced is transmitted via the microphone 7 to the voice recognition section 6 which recognizes the voice and the volume, and here it is compared with recognition data and standard volume level data and the result of this comparison are sent to the game machine main unit 1. The game main unit 1 performs a variety of operations depending on the result. The game machine main unit 1 is capable of setting the standard volume level in the voice recognition section 6. The standard volume level also has a default value which can be adjusted.

Next, a concrete example of device operation using a combination of voice recognition and volume level is described with reference to a game machine.

### 30 (Controlling degree of action by voice level)

For example, after a "forward" command has been issued to a character on the screen, the player cries "faster, faster" to increase the character's speed of movement. Alternatively, when applying an expression, such as a tearful smile etc., to the character, then the intensity of the expression is determined by the volume of the player's voice saying "Ahhh!", or the like. Alternatively, if the character on the screen is walking through perilous terrain, then when a dangerous situation arises unexpectedly, for instance, an enemy character appears suddenly, by shouting "watch out!", "car!", or the like, the player can make the character slow down its speed of movement, adopt a defensive stance, lower its posture, fall unconscious, or the like, in response to the volume of the player's voice. Alternatively, the character can be made to give a strong kick or a weak kick, depending on the volume level.

There are several different ways of perceiving the idea of volume level. Players have various ways of speaking, but in conceptual terms, the two examples shown in Fig. 6 and Fig. 7 can be given. The vertical axis in these diagrams is the volume level, and the horizontal axis represents time. In Fig. 6, the initial volume is the loudest, and the volume gradually declines. A voice in an excited state making a sudden call such as "car!", or the like, might change in this way. In Fig. 7, however, the volume rises gradually to a peak, and then declines gradually. A normal voice might change in this way.

Volume level is defined in the following way.

#### (1) Average volume

50 The average is found for the whole signal in Fig. 6 and Fig. 7 and this is taken as the (average) volume. In specific terms, the microphone input signal should be supplied to an integrating circuit. Since the average volume is a value for the whole voice (volume x time of duration), regardless of whether or not the voice is in a normal state or an excited state, it is applied to controlling rates and the degree of simple actions, such as the speed of movement of the character.

#### 55 (2) Peak volume

The level  $\Delta V$  of the signal peak value (point P in Fig. 6 and Fig. 7) is determined and this is taken as the peak volume. In specific terms, the maximum value is held continuously whilst the signal is sampled using a sample hold circuit.

The peak volume represents the maximum volume, and therefore it is suitable for use in cases where there is competition in voice level between a plurality of players.

(3) Rise time

5 The time taken to reach the signal peak ( $\Delta t$  in Fig. 6 and Fig. 7) is determined and this is taken as the rise time. This can be found easily from the hold timing of the peak hold circuit. Alternatively, the rise time can be defined by means of the rise angle  $\theta$ , and  $\Delta y/\Delta t$ . This can be achieved by means of a differentiating circuit. Since the rise time can be regarded as corresponding to the urgency of the voice, then the character can be made to adopt a defensive stance  
10 or take retreating action when the rise is sudden, and the character can be made to walk quickly, or the like, when the rise is not sudden.

(4) Continuation time

15 The time of duration of the signal ( $T$  in Fig. 6 and Fig. 7) is determined, and this is taken as the continuation time. Specifically, the time for which the voice signal continues at or above a predetermined threshold value is measured using a timer (omitted from diagram). The continuation time can be applied readily to cases where there is competition between the length of voices.

20 (Controlling type of action by voice level)

In addition to the degree of action, the type of action may also be changed by means of the voice level. For example, after a "forward" command has been issued to a character on the screen, if the player says "go" in a soft voice, the character starts to walk, whereas if he or she says "GO!" in a loud voice, the character may Jump, or if it has a flying capability, it will fly through the sky.  
25 In brief, the device can be controlled such that the degree of an action is increased or reduced by the voice level, or the type of action is changed depending on whether or not the voice level exceeds a certain threshold value.

(Controlling the type of action by the type of voice)

30 A distinction can be made between a child's voice and an adult's voice, or between a male voice and a female voice, and the action can be changed on the basis of this result. For example, if a child character and an adult character are present together, the child character or adult character can be controlled individually depending on the type of voice. This applies similarly to a male and female character. By controlling the device in this way, the game can be played even  
35 when a plurality of players speak simultaneously, and hence the game can be broadened in scope.

Incidentally, a child's voice and an adult voice can be identified by the difference in their frequency bands. A male voice and a female voice can also be identified by similar means by the difference in their frequency bands.

(Other)

40 Special actions can also be implemented by a combination of voice level and/or type of voice, and operation of the buttons on the control pad 2b. For example, the game may be controlled by a combination of a raised voice + operation of button A such that the same move is repeated. Also, in a fighting game where a deathblow is performed when a plurality of buttons are pressed simultaneously, if voice input is assigned to a certain button and normal pad input is  
45 assigned to other buttons, then it is possible to effect a similar key input by a more simple method using this combination of voice and pad input, compared to key input using a pad only.

If the voice level is used as described above, the aforementioned real-time volume output function may be used. In other words, the voice recognition mode and real-time volume output mode are combined. Firstly, after carrying out voice recognition and implementing the corresponding action, the device enters real-time volume output mode, and outputs the voice level only. This means that the response is extremely good.

50 As described above, according to this mode for implementing the invention, it is possible to control the degree of an action on the basis of the voice volume level, such that a game character makes a small movement when the player speaks a word softly, and it makes a large movement when the player speaks a word loudly. Moreover, it is also possible to control the type of action on the basis of the voice volume level. The type of action can also be controlled on the basis  
55 of the type of voice.

**Second mode for Implementing this Invention**

The device according to the first mode for implementing this invention described above adds depth to the control of the device by combining a voice recognition device with a conventional control device in the form of a pad. However, it is also possible to replace the game pad with a voice recognition device.

5 The voice recognition section 6 in Fig. 3 has a default non-specific speaker voice recognition function and a normal non-specific speaker voice recognition function. Therefore, since prescribed voice sounds correspond to pad buttons even in the default state immediately after switching on the power, it is possible to use a voice recognition device even with software which is not compatible with voice recognition. Furthermore, by issuing a dictionary data load command  
10 for the normal non-specific speaker voice recognition function to download a dictionary from a CD-ROM, it is possible to make the relationship between the pad buttons and voice recognition results compatible with the particular software in use.

Next, the processing implemented when a normal non-specific speaker voice recognition function is used will be described.

15 Firstly, dictionary data is downloaded from a CD-ROM by means of the voice recognition section 6 issuing a dictionary data load command.

Next, the correspondence between the pad 2b buttons and prescribed voice sounds is determined on the basis of the dictionary data. Firstly, voice sounds corresponding to the main unit key switches are defined in the voice data RAM, such that, for example, when the voice sound "forward" is input, key data indicating "up" is output. In this state, when  
20 the sound "forward" is input via a microphone as configured in Fig. 1, this data is read into a voice comparing section in the form of digital data, and it is then determined whether or not the same data is present in the voice data RAM. If the data which is the same as the sound input by the microphone is present in the voice data RAM, key data indicating "up" is sent to the game machine main unit 1 in accordance with this data.

When a voice signal is input, the voice recognition section 6 recognizes the voice sound and outputs key data corresponding to normal game pad operation to the game machine main unit 1 on the basis of the correspondence relationships described above. For example, if the player says "start game", then the same signal is output as when the start button is pressed. It is also possible for the player to record his or her preferred voice sounds for each button on the pad. For example, a special command for this purpose is input and then a desired word is selected or input and recorded for each item of pad data. If this is done by selecting text of the TV screen, then non-specific speaker voice recognition is possible, and if desired words are input via the microphone 7, then specific speaker voice recognition is possible (catch-up function).

In conventional data input devices for game machines, data has been input to the game machine by pressing switches provided on a game pad using the hands, fingers, or the like, but in the method described above, a game can be controlled in the same manner by inputting voice sounds.

35 Furthermore, by using a voice recognition section 6, it is possible to play the game differently to when using a game pad. In a fighting game, for instance, normally, highly skilled pad operation is required to produce a very difficult manoeuvre, but when a voice recognition section 6 is used, the game can be devised such that this manoeuvre is not produced unless a high level voice input (rapid speech or difficult pronunciation, etc.) is received. It is also possible to combine recognition of voice volume level and voice type.

40 By using voice sounds to perform key inputs in this way, it is possible to make the game feel more intimate. For example, in a boxing game where button A is assigned to a left-hand punch and button B is assigned to a right-hand punch, the game will be more exciting if punches are thrown by means of the voice sounds "left" and "right", rather than using buttons A and B.

**45 Third mode for Implementing this Invention**

After voice recognition, the voice recognition section 6 outputs a recognized word number and a similarity level. If the rejection level is set when the recognition start command is issued, then it is possible to prevent a recognition result from being output when the similarity level for that result is lower than a set level. By means of this function, it is possible  
50 to prevent the voice of someone other than the speaker, or external noise, etc. from being recognized in error.

As well as preventing incorrect recognition, the rejection level can also be applied in the following ways:

**Changing rejection level according to game circumstances**

55 In an adventure game, for example, in a scene where the player's character may die and the game will be over if there is an incorrect recognition of some kind, the rejection level can be set to a high level. In this case, the reliability of the output results is raised at the expense of the response time.

On the other hand, by lowering the rejection level in scenes which are not so important to the progress of the game,

the reliability of the output results is reduced, but the response time is improved.

Since the game developers can alter these settings as desired, a greater flexibility of control is possible compared to when game pads are used.

#### 5    Fourth mode for implementing this invention

After voice recognition, the voice recognition section 6 outputs a recognized word number and a similarity level. Therefore, the degree of an action and the type of action can be controlled according to the similarity level.

For example, if a certain word is recognized, the response can be altered depending on the size of the similarity level output by the voice recognition section 6. For example, the force in a punch resulting from an 80 point similarity level will be strong, whilst the force in a punch having a 10 point similarity level will be weak. In this way, the force of a punch can be varied according to the similarity level. In this case, if non-specific speaker voice recognition is applied, the actions of the characters or the development of the game can be changed depending on the player's pronunciation, it being more advantageous if the player's voice conforms to the standard data. The game can also be enjoyed from this point of view (namely, who has the most standard pronunciation? Whose pronunciation is suitable for voice recognition? etc.) Therefore, the players' enjoyment of the game is further enhanced.

This method of control based on the similarity level according to the fourth mode of implementation may also be applied to the specific example described in the first mode of implementing this invention in relation to the voice level.

#### 20    Fifth mode of Implementing the Invention

The description in the foregoing embodiments centred on the voice recognition section, but this mode of implementation relates to a game which is played using this voice recognition section.

##### 25    (1) Action game for animal training assistance

In this type of game, the player can communicate with animals, such as a dog, cat, horse, or the like, using voice recognition.

Voice recognition can be used to control not just a single character, but a plurality of characters, directly or indirectly. Controlling this type of game by buttons is relatively unappealing, and control by voice, similar to a real-life situation, is more suitable.

Furthermore, in real animal training, the animals sometimes ignore a person's commands. So even if voice recognition does not work perfectly, the impatience that this induces actually becomes a feature of the game, and therefore the appeal of the game is further enhanced. The variety of responses of the character when it is called to is interesting.

Alternatively, an action game can also be envisaged where an animal such as monkey, pheasant, dog, or the like, accompanying the player can be made to move around as the player wishes. In a game based on the story of *Momotaro*, for instance, the player can befriend animals in a field by actually talking. For example, in order to befriend an animal, the player calls out something nice when the animal he or she wishes to befriend is looking at the player's character. For example, if the animal is a dog, then it will respond if the player calls out "Here boy! "Here boy!", or the like, a number of times. If the animal is then given some food, it will become the player's companion.

If the player walks somewhere, the animal will follow afterwards, but if it sees something interesting, it will stop and try to go off in that direction. In this case, the animal can be brought back by calling out "Here boy! NOW!", or the like.

Animals always behave in a variety of ways. If the player repeats the same words a number of times for an action involving an animal, the animal will always perform the action corresponding to those words. By means of this repetition, the player can gradually control the animal.

Alternatively, a game where animals clear traps or enemies may also be conceived. This type of game may involve a shepherd leading sheep, for instance.

In a game of this kind, a variety of traps are prepared and the player controls the animals so that they avoid the traps. Animals can get caught in traps that are identified easily by the player, so the player must instruct the animals. As the number of friendly animals increases, the game becomes more difficult. In order to clear the traps and progress onwards, the player must consider how to handle different types of animal.

Alternative games include a tropical fish breeding game, a golden carp breeding game, or the like. The fish can be gathered together not only by voice control but also by an action, such as clapping hands. By programming the device such that the fish perform a variety of actions depending on how the player claps his or her hands, an interesting way of playing the game can be provided.

## (2) Multiple player rapid response quiz

This relates to a quiz program in a game machine. In this game, there are a plurality of participants and the game machine puts questions to the players in turn, and the players compete to get the highest score. The response words 5 are limited to (1), (2), ... or (a), (b), ... etc., which is advantageous in terms of processing speed.

## (3) Children's educational game

A children's character is made to remember or practice words. The object of this type of game is communication, 10 and it is made more interesting if voice control is used.

## (4) Flight simulation game

Using a voice recognition section for selecting weapons and other equipment, or for giving commands to a co-pilot, 15 makes the game more realistic.

## (5) Mahjong game

The player can choose hands by making Mahjong calls. 20

## (6) Other

Other conceivable applications include: *shogi* (Japanese chess), table games, party games (conversation-based games, rather than mouse-based games), number games (*kanojo*, *ocha-shinai*, *OK*, etc.), driving games (radio control), RPG (role playing game) short cuts, conversation-based virtual family games (for elderly people), and war simulation games, etc. 25

A recording medium is a CD-ROM, cartridge, or the like, on which information (principally digital data and programs) is recorded by some physical means or other, and it is able to cause a processing device, such as a dedicated processor, or the like, to implement prescribed functions. In brief, it should download programs to a computer by some 30 means of other and cause the computer to execute prescribed functions.

Examples of this include floppy disks, hard disks, magnetic tape, magneto-optical disks, CD-ROM, DVD, ROM cartridges, RAM memory cartridges equipped with back-up batteries, flash memory cartridges and non-volatile RAM cartridges.

Wired communications circuits, such as public telephone lines, radio communications circuits, such as microwave 35 circuits, and communications media such as the Internet are also included.

INDUSTRIAL APPLICABILITY

In a voice recognition device used as a peripheral device in a game machine, since the voice recognition device 40 relating to this invention comprises voice input means, and a voice recognition section for recognizing the player's voice by comparing the voice signal output from this voice input means with data from a previously defined voice recognition dictionary, and generating control signals relating to the game on the basis of the recognition result, a person playing a game can control the game using his or her voice

In the voice recognition device relating to this invention, since the voice recognition section comprises a non-specific 45 speaker voice recognition dictionary, which is previously defined for unspecified speakers, and a specific speaker voice recognition dictionary which is defined by the player, and in its initial state, the device selects the non-specific speaker voice recognition dictionary, a game based on voice control can be played by means of non-specific speaker voice recognition, even for game software which is not compatible with specific speakers. This is not limited to specialized software, but can be applied to many different types of software, thus broadening the scope of application of the 50 device.

In the voice recognition device relating to this invention, since the voice recognition section comprises a plurality of specific speaker voice recognition dictionaries corresponding respectively to a plurality of players, and one of these specific speaker voice recognition dictionaries is selected by an action of the player as the dictionary to be used for voice 55 recognition processing, it is possible to apply specific speaker voice recognition using an appropriate dictionary, and the accuracy of voice recognition can be raised.

Since the voice recognition device relating to this invention comprises a game machine control section connected to the voice recognition section, and the voice recognition section generates control signals relating to the game by combining voice recognition results and control signals from the control section, a variety of actions are possible by

combining a normal game pad, or the like, and voice recognition, and thus games can be made more interesting. In the voice recognition device relating to this invention, since the control section outputs control signals for implementing normal actions, and the voice recognition section generates control signals for implementing special actions, special actions, which are difficult to execute using a normal game pad etc., can be performed more easily.

- 5      In the voice recognition device relating to this invention, since the voice recognition section outputs a value indicating the state of the voice signal and a similarity level indicating the degree of similarity between the voice signal output from the voice input means and the contents of the voice recognition dictionary, and the voice recognition section has a first operating mode, wherein voice recognition is conducted on the basis of the similarity level in order to select the type of game action, and a second operating mode, wherein no voice recognition is conducted and the state of the voice
- 10     signal is measured in order to set the state of the action, the processing in the second operating mode can be conducted at high speed, and the response is fast, which is beneficial in terms of controlling the game.

In the voice recognition device relating to this invention, since the voice recognition section outputs a similarity level indicating the degree of similarity between the voice signal output from the voice input means and the contents of the voice recognition dictionary, and the corresponding volume level, evaluates this volume level on the basis of a predetermined rejection level, and rejects the recognition result from the voice recognition section depending on this evaluation result, it is possible to avoid voice recognition when an appropriate voice recognition result is not obtainable, and hence recognition errors can be reduced.

Since the voice recognition device relating to this invention sets a rejection level for each type of game or each game stage, an appropriate rejection level can be set for each scene, which is advantageous in playing a game.

- 20     In the voice recognition device relating to this invention, since the voice recognition section performs voice recognition on the basis of the similarity level indicating the degree of similarity between the voice signal output from the voice input means and the contents of the voice recognition dictionary, and changes the state of action in the game according to control signals generated on the basis of this recognition result, in response to this similarity level, special actions, which are difficult to execute using a normal game pad, or the like, can be performed more easily.

## 25     Claims

1. A voice recognition device used as a peripheral device for a game machine, a voice recognition device comprising:  
  - 30     voice input means; and  
 a voice recognition section for recognizing the player's voice by comparing the voice signal output from said voice input means with data from previously defined voice recognition dictionaries, and generating control signals relating to the game on the basis of the recognition result.
2. A voice recognition device according to claim 1, wherein said voice recognition section comprises, as said voice recognition dictionaries, a non-specific speaker voice recognition dictionary, which is previously defined for unspecified speakers, and a specific speaker voice recognition dictionary which is defined by the player, and in its initial state, the device selects said non-specific speaker voice recognition dictionary.
3. A voice recognition device according to claim 2, wherein said voice recognition section comprises a plurality of specific speaker voice recognition dictionaries corresponding respectively to a plurality of players, and one of these specific speaker voice recognition dictionaries is selected by an action of the player as the dictionary to be used for voice recognition processing.
4. A voice recognition device according to claim 1, comprising a game machine control section connected to said voice recognition section, and said voice recognition section generates control signals relating to the game by combining voice recognition results and control signals from said control section.
5. A voice recognition device according to claim 3, wherein said control section outputs control signals for implementing normal actions, and said voice recognition section generates control signals for implementing special actions.
6. A voice recognition device according to claim 1, wherein said voice recognition section outputs a value indicating the state of the voice signal output from said voice input means and a similarity level indicating the degree of similarity between said voice signal and the contents of said voice recognition dictionary, and said voice recognition section has a first operating mode, wherein voice recognition is conducted on the basis of the similarity level in order to select the type of game action, and a second operating mode, wherein no voice recognition is conducted and the state of said voice signal is measured in order to set the state of the action.

7. A voice recognition device according to claim 6, wherein, in said second mode, takes the average sound level of a part or all of the voice signal as the state of said voice signal, and sets the state of the game action on the basis of this average sound level.
  
- 5 8. A voice recognition device according to claim 6, wherein, in said second mode, takes the peak sound level of said voice signal as the state of the voice signal, and sets the state of the game action on the basis of this peak sound level.
  
9. A voice recognition device according to claim 6, wherein, in said second mode, takes the voice signal rise time as  
10 the state of said voice signal, and sets the state of the game action on the basis of this rise time.
  
10. A voice recognition device according to claim 6, wherein, in said second mode, takes the voice signal continuation time as the state of said voice signal, and sets the state of the game action on the basis of this continuation time.
  
- 15 11. A voice recognition device according to claim 6, wherein, in said second mode, takes the type of voice as the state of said voice signal, and sets the state of the game action on the basis of this voice type.
  
12. A voice recognition device according to claim 1, wherein said voice recognition section outputs a similarity level indicating the degree of similarity between the voice signal output from said voice input means and the contents of  
20 said voice recognition dictionary, and the corresponding volume level, evaluates this volume level on the basis of a predetermined rejection level, and rejects the recognition result from said voice recognition section depending on this evaluation result.
  
- 25 13. A voice recognition device according to claim 12, sets said rejection level for each type of game or each game stage.
  
14. A voice recognition device according to claim 1, wherein said voice recognition section performs voice recognition on the basis of the similarity level indicating the degree of similarity between said voice signal output from the voice input means and the contents of said voice recognition dictionary, and changes the state of action in the game according to control signals generated on the basis of the recognition result, in response to said similarity level.  
30
  
15. A game machine comprising a voice recognition device according to any of claims 1 to 14 as a peripheral control device.
  
- 35 16. A voice recognition method, comprising:
  - a first step whereby a voice signal is received;
  - a second step whereby a player's voice is recognized by comparing said voice signal with data from a previously defined voice recognition dictionary; and
  - 40 a third step whereby control signals relating to the game are generated on the basis of the recognition result from said second step.
  
17. A voice recognition method according to claim 16, comprising a fourth step whereby control signals are received from the game machine control section, wherein, in said third step, control signals relating to the game are generated by combining the voice recognition result from said second step and the control signals from said fourth step.  
45
  
18. A medium for recording programs for executing voice recognition operations in said voice recognition section.

## Amended claims under Art. 19.1 PCT

50

1. (Delete)
  
2. (Amended) A voice recognition device used as a peripheral device for a game machine:
  - 55 voice input means; and
  - a voice recognition section for recognizing the player's voice by comparing the voice signal output from said voice input means with data from previously defined voice recognition dictionaries, and generating control signals relating to the game on the basis of the recognition result,

wherein said voice recognition section comprises, as said voice recognition dictionaries, a non-specific speaker voice recognition dictionary, which is previously defined for unspecified speakers, and a specific speaker voice recognition dictionary which is defined by the player, and in its initial state, the device selects said non-specific speaker voice recognition dictionary.

- 5        3. A voice recognition device according to claim 2, wherein said voice recognition section comprises a plurality of specific speaker voice recognition dictionaries corresponding respectively to a plurality of players, and one of these specific speaker voice recognition dictionaries is selected by an action of the player as the dictionary to be used for voice recognition processing.
- 10      4. (Amended) A voice recognition device according to claim 2, comprising a game machine control section connected to said voice recognition section, and said voice recognition section generates control signals relating to the game by combining voice recognition results and control signals from said control section.
- 15      5. A voice recognition device according to claim 3, wherein said control section outputs control signals for implementing normal actions, and said voice recognition section generates control signals for implementing special actions.
- 20      6. (Amended) A voice recognition device according to claim 2, wherein said voice recognition section outputs a value indicating the state of the voice signal output from said voice input means and a similarity level indicating the degree of similarity between said voice signal and the contents of said voice recognition dictionary, and said voice recognition section has a first operating mode, wherein voice recognition is conducted on the basis of the similarity level in order to select the type of game action, and a second operating mode, wherein no voice recognition is conducted and the state of said voice signal is measured in order to set the state of the action.
- 25      7. A voice recognition device according to claim 6, wherein, in said second mode, takes the average sound level of a part or all of the voice signal as the state of said voice signal, and sets the state of the game action on the basis of this average sound level.
- 30      8. A voice recognition device according to claim 6, wherein, in said second mode, takes the peak sound level of said voice signal as the state of the voice signal, and sets the state of the game action on the basis of this peak sound level.
- 35      9. A voice recognition device according to claim 6, wherein, in said second mode, takes the voice signal rise time as the state of said voice signal, and sets the state of the game action on the basis of this rise time.
- 40      10. A voice recognition device according to claim 6, wherein, in said second mode, takes the voice signal continuation time as the state of said voice signal, and sets the state of the game action on the basis of this continuation time.
- 45      11. A voice recognition device according to claim 6, wherein, in said second mode, takes the type of voice as the state of said voice signal, and sets the state of the game action on the basis of this voice type.
- 50      12. (Amended) A voice recognition device according to claim 2, wherein said voice recognition section outputs a similarity level indicating the degree of similarity between the voice signal output from said voice input means and the contents of said voice recognition dictionary, and the corresponding volume level, evaluates this volume level on the basis of a predetermined rejection level, and rejects the recognition result from said voice recognition section depending on this evaluation result.
- 55      13. A voice recognition device according to claim 12, sets said rejection level for each type of game or each game stage.
- 60      14. (Amended) A voice recognition device according to claim 2, wherein said voice recognition section performs voice recognition on the basis of the similarity level indicating the degree of similarity between said voice signal output from the voice input means and the contents of said voice recognition dictionary, and changes the state of action in the game according to control signals generated on the basis of the recognition result, in response to said similarity level.

15. (Amended) A game machine comprising a voice recognition device according to any of claims 2 to 14 as a peripheral control device.

16. A voice recognition method, comprising:

5  
a first step whereby a voice signal is received;  
a second step whereby a player's voice is recognized by comparing said voice signal with data from a previously defined voice recognition dictionary; and  
10 a third step whereby control signals relating to the game are generated on the basis of the recognition result from said second step.

17. A voice recognition method according to claim 16, comprising a fourth step whereby control signals are received from the game machine control section, wherein, in said third step, control signals relating to the game are generated by combining the voice recognition result from said second step and the control signals from said fourth step.

15  
18. A medium for recording programs for executing voice recognition operations in said voice recognition section.

20

25

30

35

40

45

50

55

FIG. 1

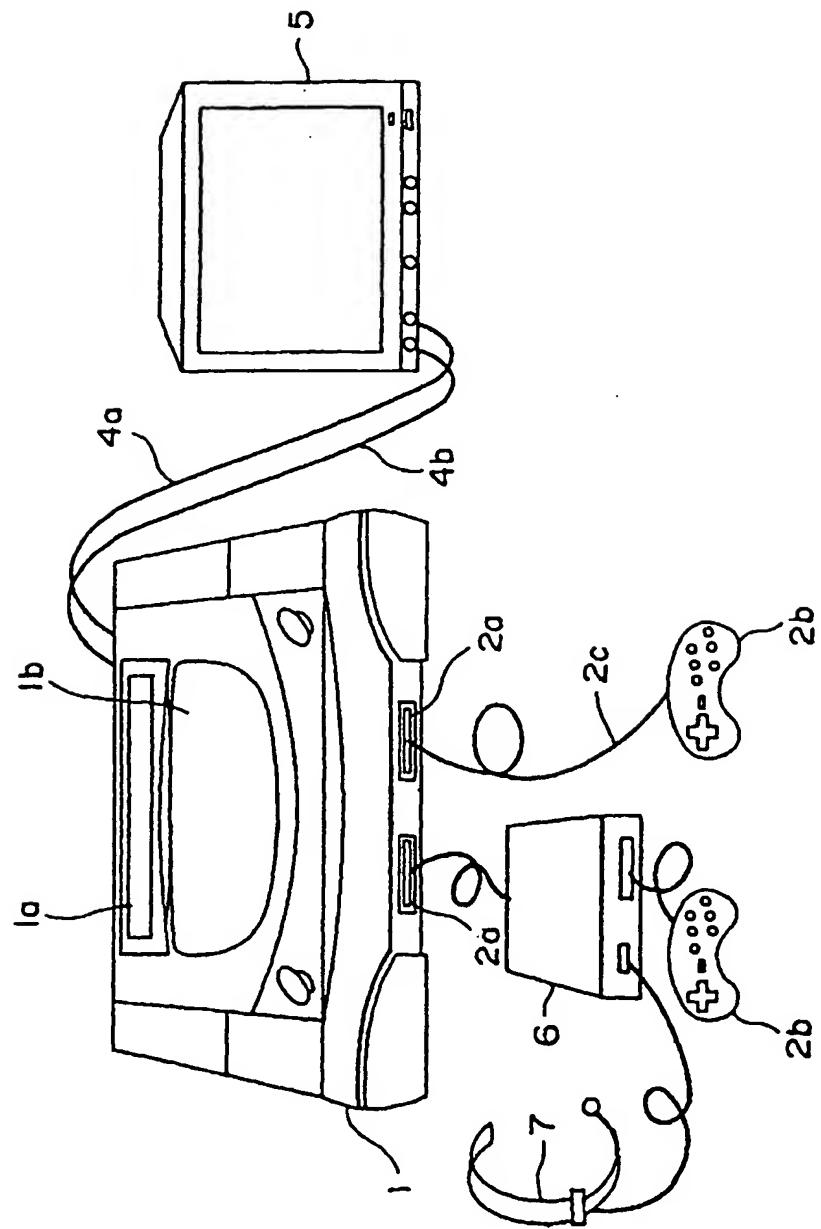


FIG. 2

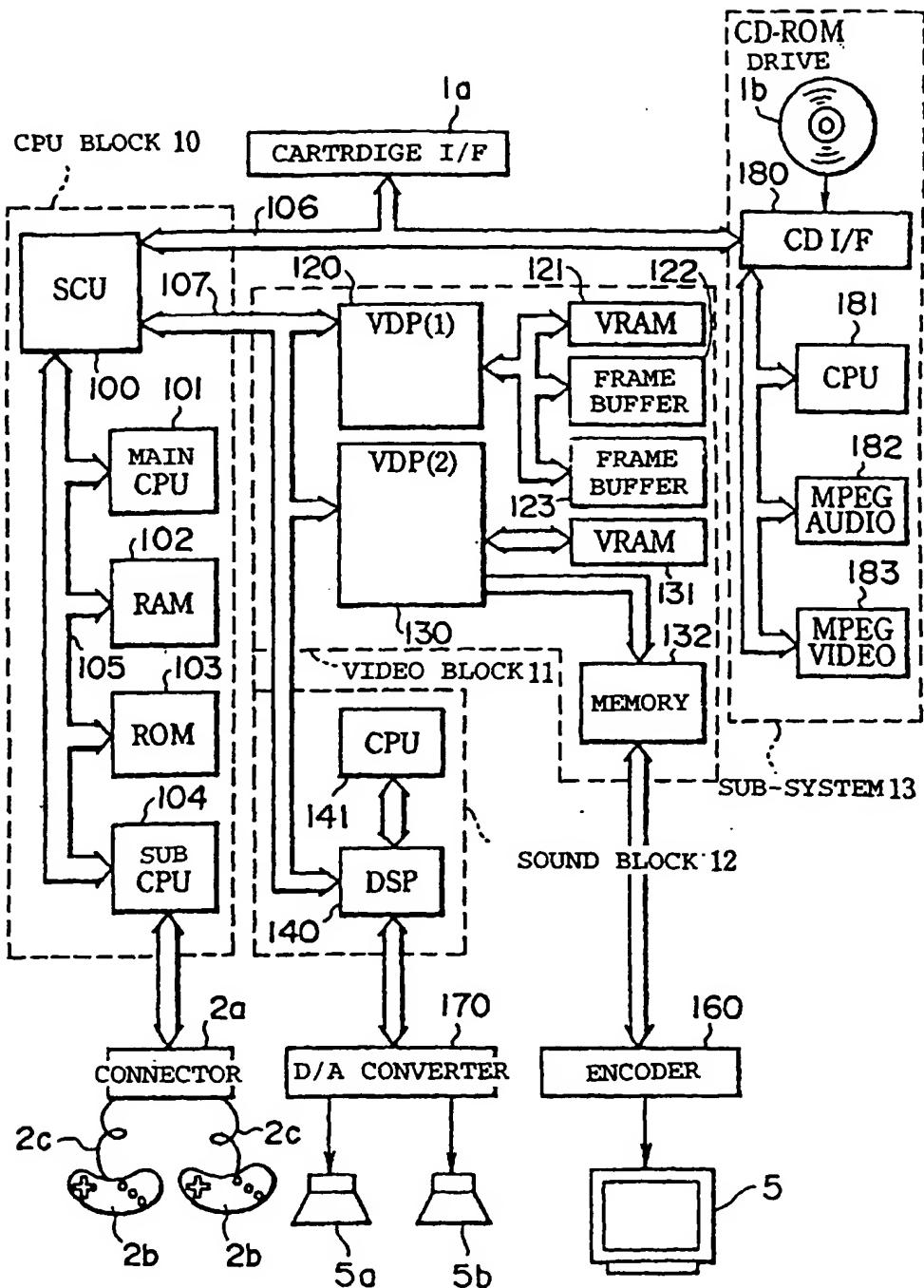


FIG. 3

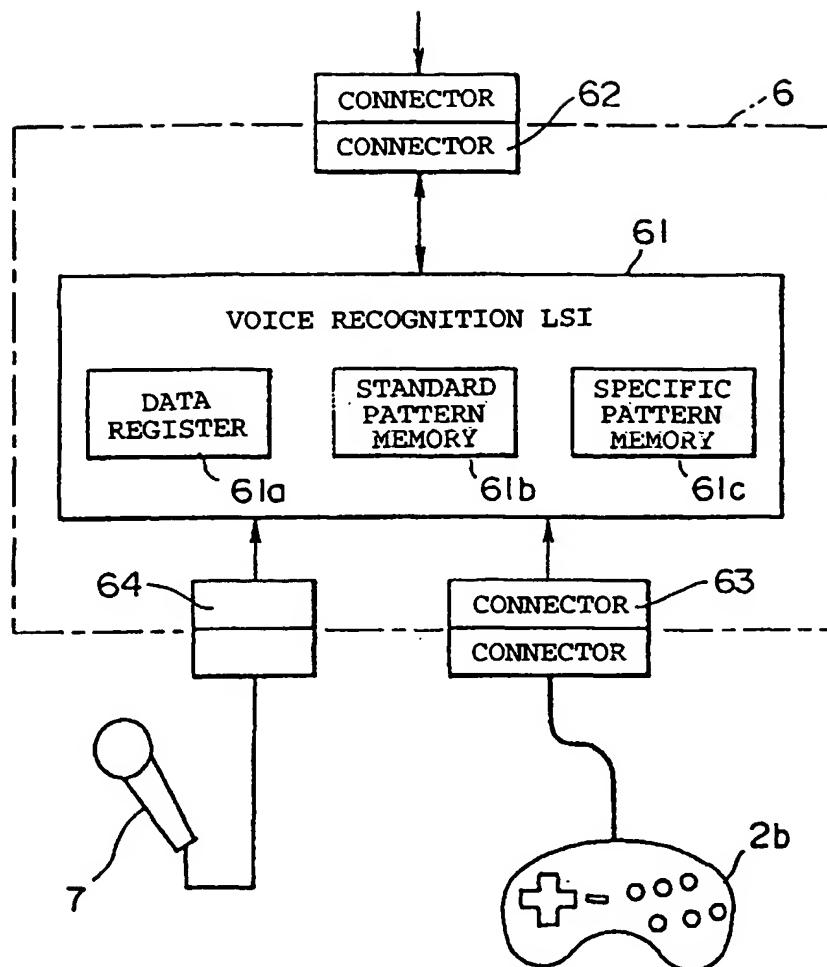


FIG. 4

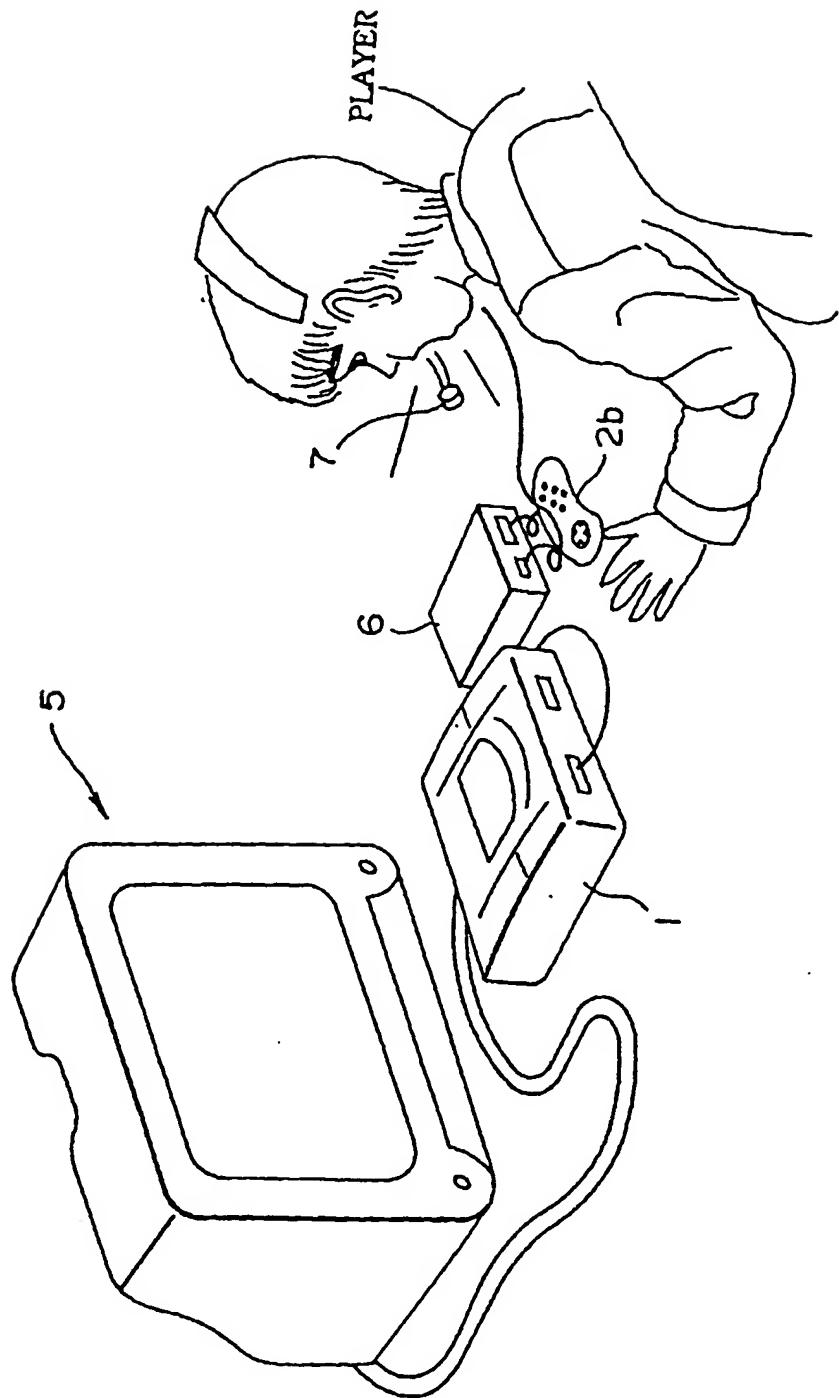


FIG. 5

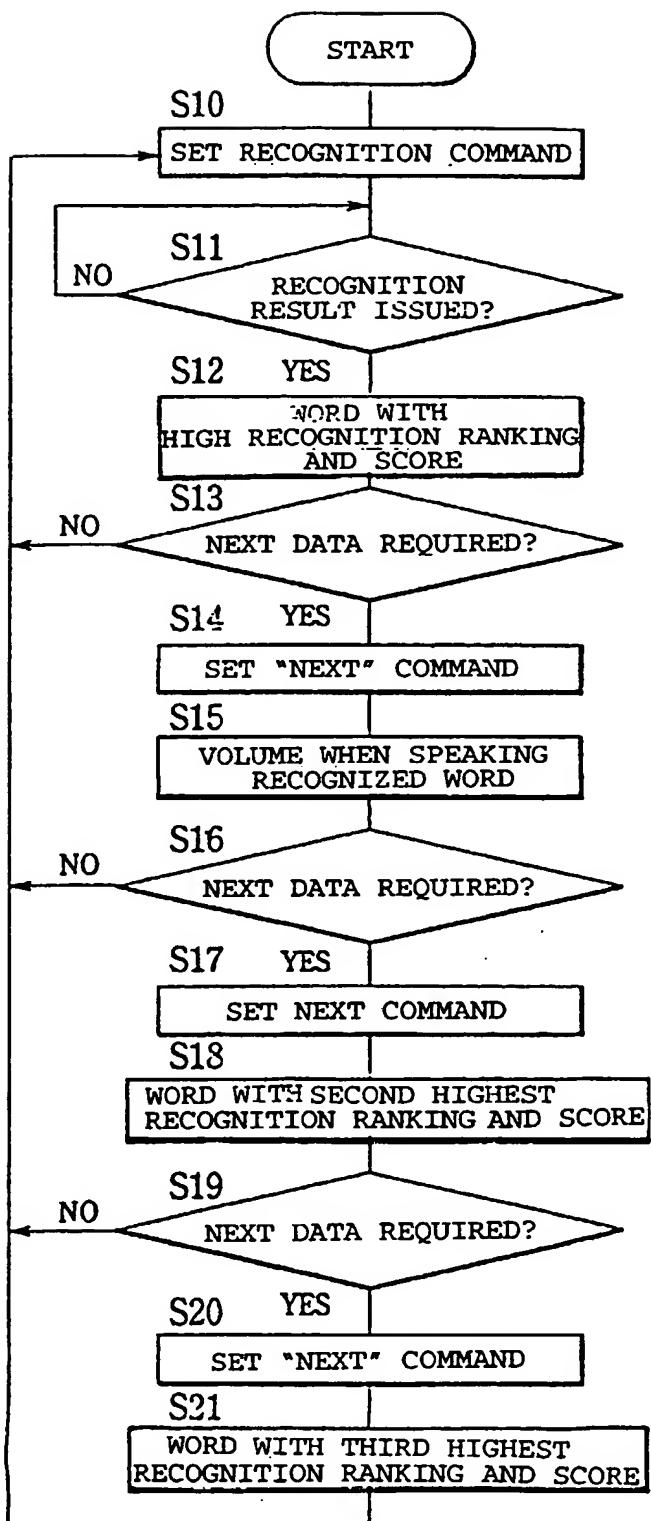


Fig. 6

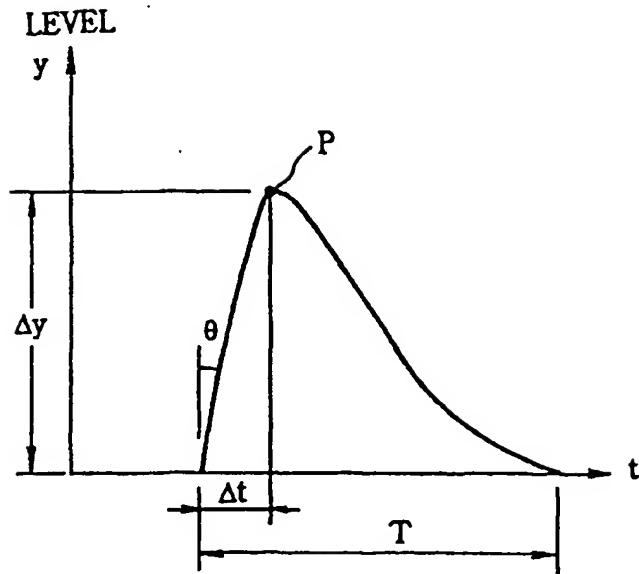
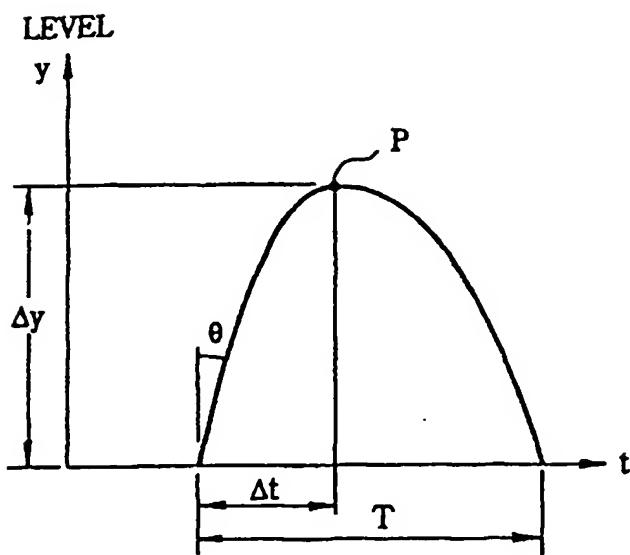


Fig. 7



INTERNATIONAL SEARCH REPORT		International application No. PCT/JP97/02421
<b>A. CLASSIFICATION OF SUBJECT MATTER</b>		
Int. Cl <sup>6</sup> A63F9/22		
According to International Patent Classification (IPC) or to both national classification and IPC		
<b>B. FIELDS SEARCHED</b>		
Minimum documentation searched (classification system followed by classification symbols)		
Int. Cl <sup>6</sup> A63F9/22, G10L3/00		
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched Jitsuyo Shinan Koho 1922 - 1996 Jitsuyo Shinan Toroku Kokai Jitsuyo Shinan Koho 1971 - 1997 Koho 1996 - 1997 Toroku Jitsuyo Shinan Koho 1994 - 1997		
Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)		
<b>C. DOCUMENTS CONSIDERED TO BE RELEVANT</b>		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	JP, 6-315572, A (Ricoh Co., Ltd.), November 15, 1994 (15. 11. 94) (Family: none) Full descriptions; Figs. 1 to 3	1, 4-5, 15-18
Y	Full descriptions; Figs. 1 to 3	2 - 3
A	Full descriptions; Figs. 1 to 3	6 - 14
A	JP, 64-23299, A (Ricoh Co., Ltd.), January 25, 1989 (25. 01. 89) (Family: none) Full descriptions; Figs. 1 to 4	1 - 18
A	JP, 4-276280, A (Sony Corp.), October 1, 1992 (01. 10. 92) (Family: none) Full descriptions; Figs. 1 to 3	1 - 18
A	Microfilm of the specification and drawings annexed to the written application of Japanese Utility Model Application No. 91363/1988 (Laid-open No. 13587/1990) (Takehiko Kikuchi), January 29, 1990 (29. 01. 90) (Family: none) Full descriptions; Figs. 1, 2	12 - 13
<input checked="" type="checkbox"/> Further documents are listed in the continuation of Box C. <input type="checkbox"/> See patent family annex.		
* Special categories of cited documents: "A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier document but published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reasons (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art "&" document member of the same patent family		
Date of the actual completion of the international search October 3, 1997 (03. 10. 97)	Date of mailing of the international search report October 14, 1997 (14. 10. 97)	
Name and mailing address of the ISA/ Japanese Patent Office Facsimile No.	Authorized officer Telephone No.	

## INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP97/02421

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	JP, 8-123448, A (Sega Enterprises, Ltd.), May 17, 1996 (17. 05. 96) (Family: none) Full descriptions; Figs. 1 to 9	1 - 18
A	JP, 7-222868, A (Sega Enterprises, Ltd.), August 22, 1995 (22. 08. 95) (Family: none) Full descriptions; Figs. 1 to 7	1 - 18
A	JP, 7-175567, A (Nippon Steel Corp.), July 14, 1995 (14. 07. 95) (Family: none) Full descriptions; Figs. 1 to 4	1 - 18

Form PCT/ISA/210 (continuation of second sheet) (July 1992)